Vocal Assistants for Reading News

Conversational Agents
University of Fribourg

Jennifer Swaminathan

Jennifer.Swaminathan@unine.ch

07.02.23

Supervisor:

Leonardo Angelini

Overview

- Introduction
- Literature review questions
- Methods
- Results
- Synthesis table
- Challenges
- Conclusion

Introduction

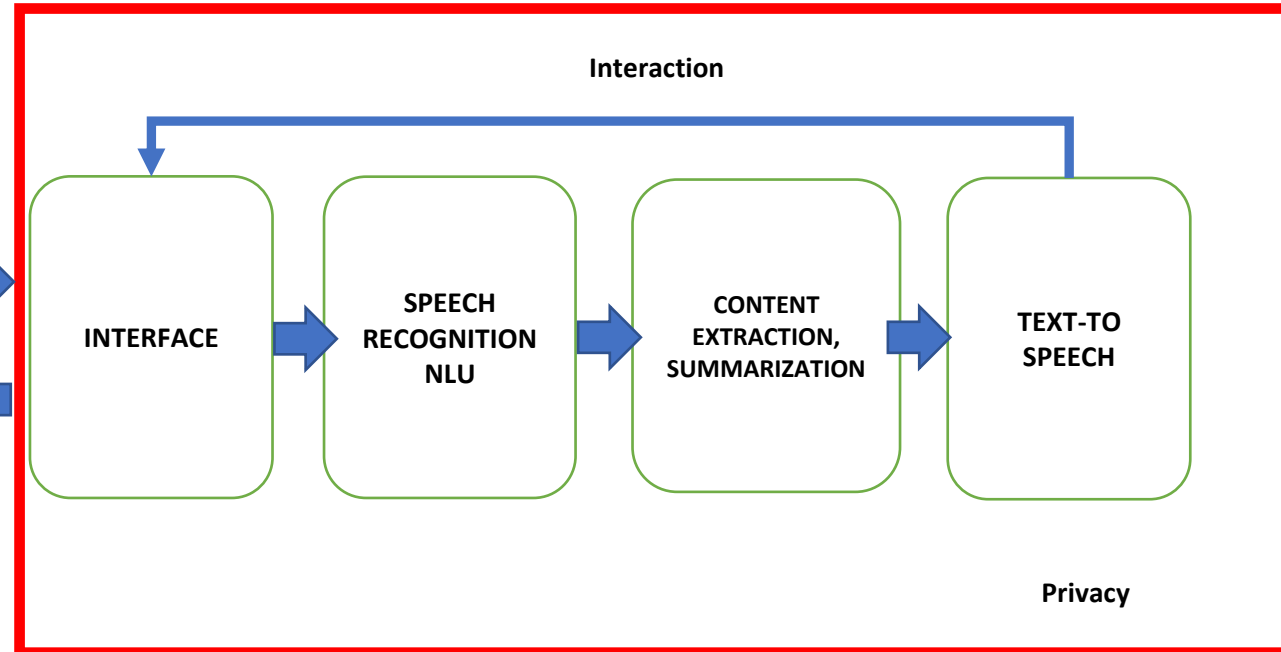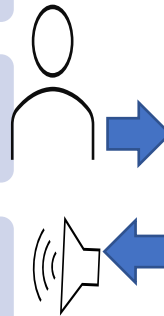Vocal Assistants

Applications

News with Vocal Assistants

Textual and vocal assistants

Pipeline

Limitations

Interaction

INTERFACE

SPEECH RECOGNITION NLU

CONTENT EXTRACTION, SUMMARIZATION
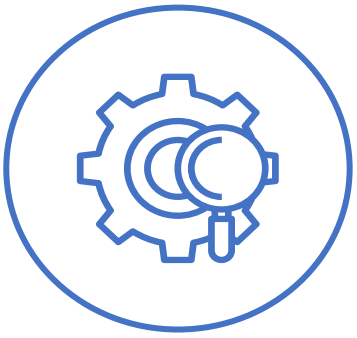
TEXT-TO SPEECH

Privacy

Research Questions:

RQ1 What are the different methods used to retain users interaction?

RQ2 Are there improvements in user experience observed with these approaches?

Methods

ACM digital library

Semantic Scholar

Vocal assistants news

Conversational agents News

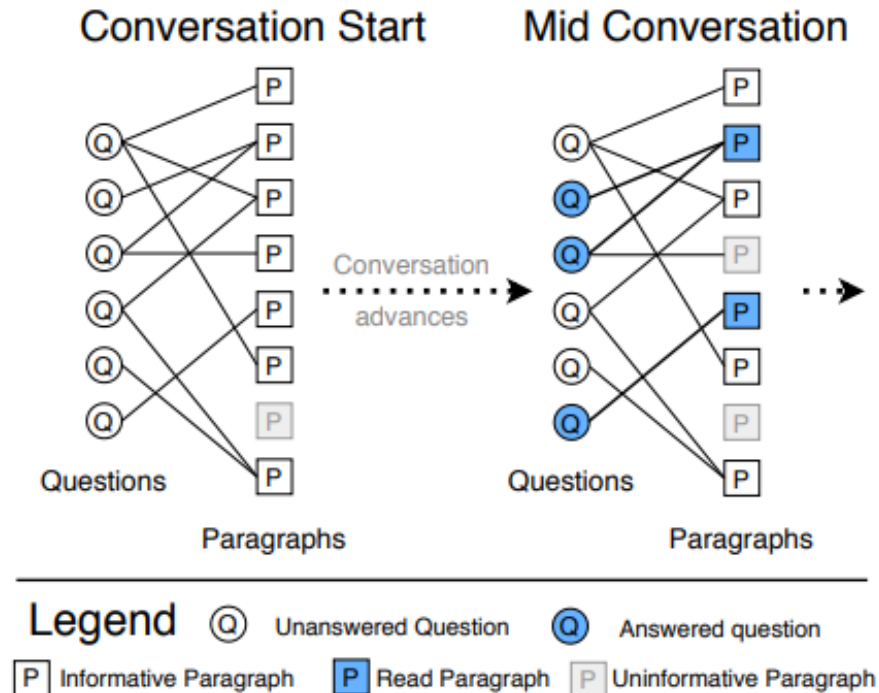News voice assistant Question Answer generation

Summarization News

15 -18 articles selected

Results

- QA generator
- Proactive vocal assistants
- Personalised vocal assistants
- Automated Podcasts and modulated voice assistants

QA Generator

- Interactivity[10]
- Non-repetitive [11]
- QGen- ACS [15]
- QAGen & DG [14]

Avoid repetition [11]

The fight scene finale between Sharon and the character played by Ali Larter, from the movie **Obsessed**, won the 2010 **MTV Movie Award for Best Fight**.

Answer: MTV Movie Award for Best Fight
Clue: from the movie Obsessed
Style: Which
Q: A fight scene from the movie, Obsessed, won which award?

Answer: MTV Movie Award for Best Fight
Clue: The flight scene finale between Sharon and the character played by Ali Larter
Style: Which
Q: Which award did the fight scene between Sharon and the role of Ali Larter win?
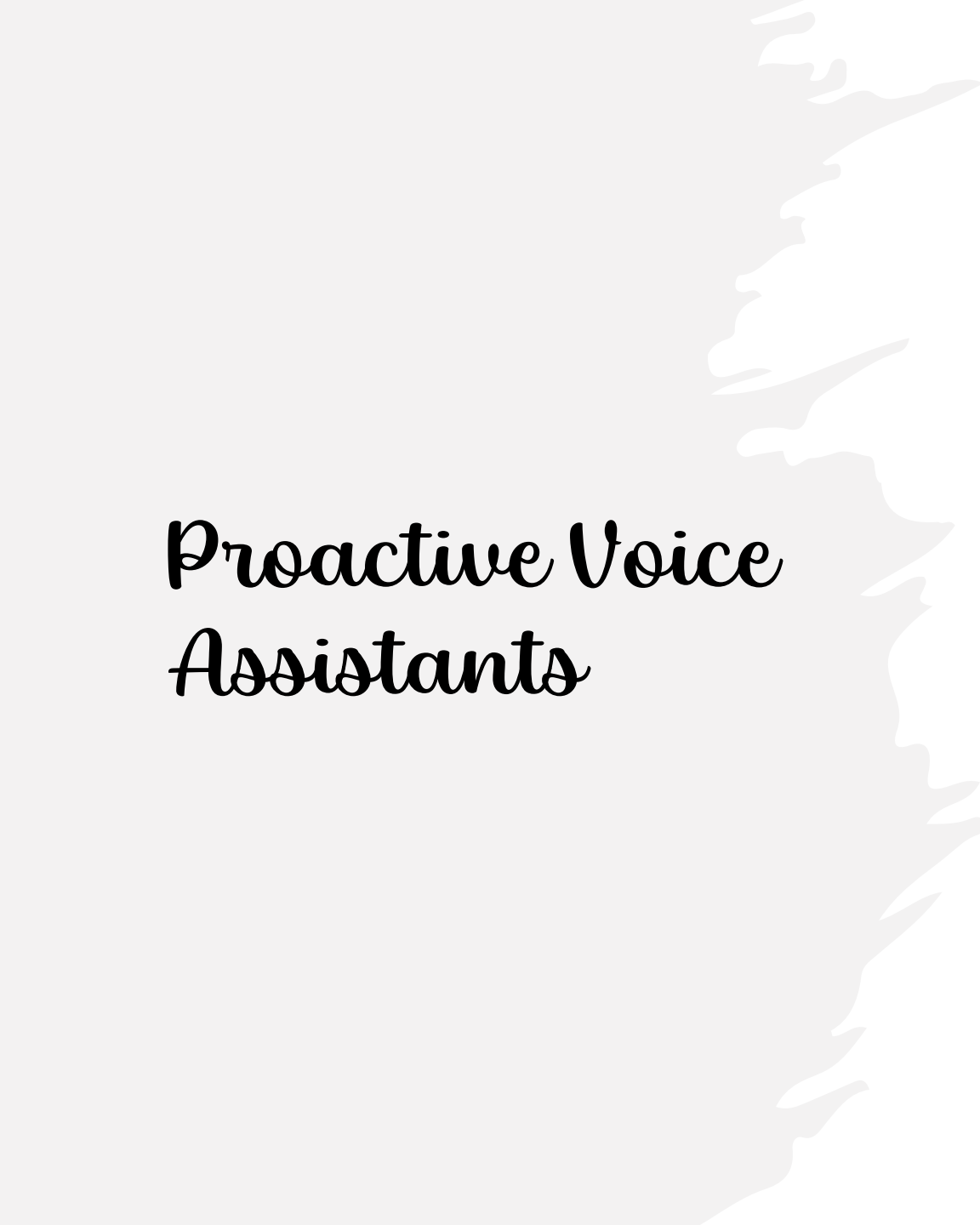
Answer: Obsessed
Clue: won the 2010 MTV Movie Award for Best Fight
Style: What
Q: What is the name of the movie that won the 2010 MTV Movie Award for Best Fight?

Qgen-ACS[15]

Proactive Voice Assistants

- Opportune moments[2]
  - Proximity
  - Context aware

- How to interrupt [19]
  - 'Hey, Are you available?'
  - Earcon starter
  - No starter

Personalised Vocal Assistants

Length of response [9]

Type of news[18]

Diversity[4]

Encourage curiosity[7]

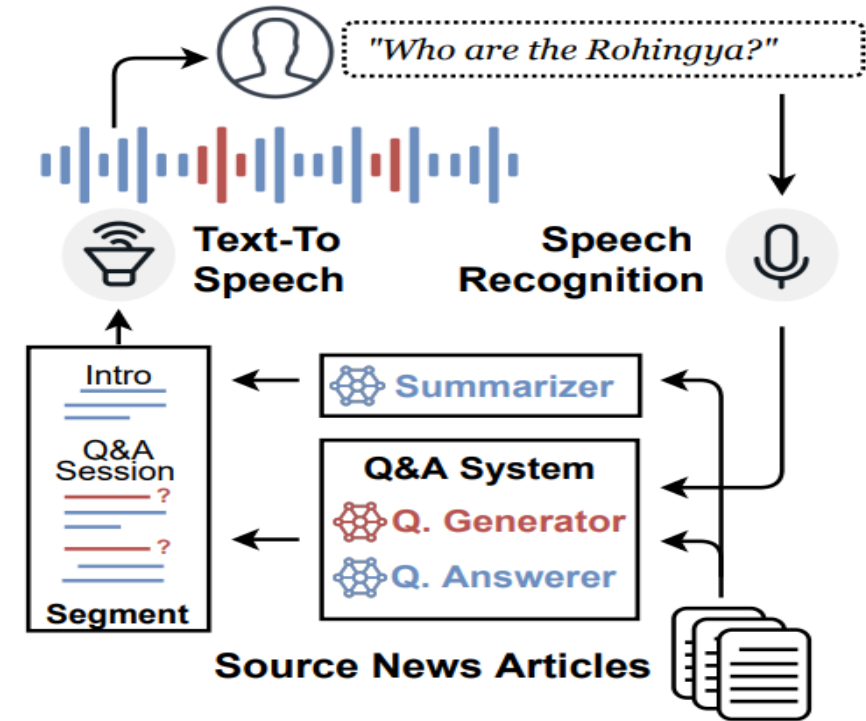| Topic	minimal	keyword	full sentence
Weather	"sunny, 5 to -3 degrees"	"Weather tomorrow: sunny, 5 to -3 degrees"	"The weather tomorrow is gonna be sunny with a high of 5 degrees and a low of -3 degrees."
Calendar	"Lunch with Steven, 12PM. Dinner with Ann, 7PM."	"Meetings tomorrow: Lunch with Steven, 12PM. Dinner with Ann, 7PM."	"You have two appointments. At 12 PM there is Lunch with Steven and at 7 PM dinner with Ann."
News	"BBC - Australia not intimidated by Facebook news ban."	"News today: BBC - Australia not intimidated by Facebook news ban."	"Here is what I found. BBCs latest headline is: Australia not intimidated by Facebook news ban."
General knowledge	"328 million."	"US population: 328 million."	"The total size of the US population amounts to 328 million."

Length of response [9]

Type	Description	Examples
Generic	Proposing to talk about news without a specific entity/headline.	"There is so much going on in the world. Would you like to talk about the news?
Trending News	Proposing to tell news on a trending topic from Bing News/Google Trends.	"How about a recent sports story titled "Astros solidify comeback". Want to hear it?" "There is a trending technology story about Apple AirPods Pro. Want to hear it?"
News Briefing	Proposing to provide a 45-second summary of the news from Reuters.	"I prepared a 45-second overview of the news for you. Want to hear it?"
Entity based (Current Conversation)	Proposing to tell news about the user's favorite sports team (if they mentioned it earlier), or any entity that the user asked a question about. (How old is the Dalai Lama?)	"Do you want to hear the latest sports updates on your favorite team?"
Entity based (Past Conversation)	Proposing to tell news about an entity mentioned by the user in past conversations.	"Last time we spoke, you mentioned Imagine Dragons. Would you like to hear the latest news on Imagine Dragons?"

Type of News [18]

Automated Podcasts & Voice modulation

- Automated Podcasts
  - Translation[1]
  - Q&A with multiple voice[12]
- Voice modulation
  - Kin voice[3]
  - Prosody modification[5]

Newspod Architecture [12]

Synthesis Table

	Tell me more[10]	whats the latest (news chatbot) [11]	Human Way Questions [15]	Quiz style- news[14]	hello there[2]	Understanding Proactive smart speakers[19]	would you like to hear news[18]	keep it short[9]	Inquistive Mind [7]	NewsPod [12]	Prosody Modification[5]	Usability Of TTS Technology in Creating News Podcast[1]	kin voice[3]
Year	2019	2021	2020	2020	2020	2021	2020	2022	2019	2022	2019	2022	2021
Aim	design a CA with interaction	Chatbot to answer any type of questions based on news	Q&A for a dataset- with varying questions like humans	Retain info with quiz format interaction	proactive interactiveto find opportune moments	note methods to interrupt	maintain engagement	VA answer with short answers	encourage curiosity: sending links and articles	Automated news podcasts	Evaluating the usefulness of audio transformation	automatic news podcasrs in different language	how users perceive the kin voice
How? Retain attention	Interaction	Avoid repetition, tracking conversation	Provides diverse question and answers	Check user is attentive with Quiz	Proactive speaker with minimal disruption	proactive speake to find interuption method	Personalised recommendation	using short answers	Increase curiosity by providing additional	SSML- for different voices	SSML for speed, rate, pitch	Using TTS to add different voices	using synthesised Kin voices
Methodology	speech synthesis API of Naver Clova	QP graph	ACS-QG	QAGenerator, distractor Generator (DG)	Sensors to monitor the users	Using sensors and voice interface	recommendation style is varied	Webspeech API and google Speech synthesis	Contrastive sentiment analysis, Natural Language	TTS API, extractive Q&A, abstractive summarization	SSML to include prosody modification	Google wavenet TTS	kinVoice- Alexa based VUI.
Advantages	interactions helps retain news	avoid repetition to retain consumer retention	Generation of questions are controlled,	A new dataset for quiz style	opportune movements found to retain attention	starter conversation should be human like	Personalised recommendation	Short response makes efficient interaction	media consumption. Serendipity in news coverage	Different voices to Q & A makes it less monotonous	Prosody modification helps in emphasing the key answers	Quick access to news	bridge physical distance
Drawbacks	participants not ready to ask questions in between	unable to handle small talk, QA is not accurate			grammatical errors were present	Camera was used to monitor the participants. Privacy concern	experiment done in dormitory.	less topics to discuss about	not customizable	Referering to too many articles could be misleading	imperfect speech & Q&A, lack of editorial judgement	Unnaturalness of the voice due to modification	Participants could identify TTS, can it be more humanlike	Synthetic kinVoice not deeveloped well
Evaluations:	Wizard of ox style evaluation	engage with multi turn questions	Quantitative and Qualitative	Case study: google platform with auto QA	a week long field study on 40 participants		Alexa users, with different types of engament	3 response styles	None	userstudies and baseline analysis	User study with TTS vs Prosody intervention	3 tests with human voice, TTS and hybrid	user study on 4 cases with kinVoice,
Results:	participants not ready to ask questions in between	people continued to have longer conversation	Quality of QG outperforms baseline	Users preferred these Quizs	Privacy convern	Utterance interruption 'are you there' was preferred.	Higher acceptance rate for entity based recommendation	short responses were preferred, but not for news	Too many resources are misleading	Less monotonous, but diverse questions	Lower speaking rate and increasing pitch points to answer	acceptance of integration of voices	For daily tasks preferred kinVoice, others eerie.
Voice	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interactivity	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	No	No
news related	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	No

11

Challenges

Voice recognition errors [6]

Reliability and accuracy of news [7]

Privacy[13]

Components: Vocal Assistants for Reading News

Interface

Speech recognition, NLU

Content extraction, Summarization

Voice and tone

Interactivity

Privacy & Security

Conclusions

Vocal assistants for reading news

Interface, Speech recognition, Content extraction, Voice&Tone, Interactivity, Privacy

Challenges:
Privacy, Source, Speech recognition

Limitations: not extensive study

References:

[1] Marcelo Carvalho Afonso, Pedro Almeida, Pedro Bec¸a, Telmo Silva, and Iulia Covalenco. Usability of text-to-speech technology in creating news podcasts using portuguese of portugal. In ACM International Conference on Interactive Media Experiences, pages 363–368, 2022

[2] Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. Hello there! is now a good time to talk? opportune moments for proactive interactions with smart speakers. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(3):1–28, 2020.

[3] Sam WT Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. Kinvoices: Using voices of friends and family in voice interfaces. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2):1–25, 2021.

[4] Li Chen, Zhirun Zhang, Xinzhi Zhang, and Lehong Zhao. A pilot study for understanding users' attitudes towards a conversational agent for news recommendation. In Proceedings of the 4th Conference on Conversational User Interfaces, pages 1–6, 2022.

[5] Aleksandr Chuklin, Aliaksei Severyn, Johanne Trippas, Enrique Alfonseca, Hanna Silen, and Damiano Spina. Prosody modifications for question-answering in voice-only settings. arXiv preprint arXiv:1806.03957, 2018.

[6] Henry Kudzanai Dambanemuya and Nicholas Diakopoulos. Auditing the information quality of news-related queries on the alexa voice assistant. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–21, 2021.

[7] Mateusz Dubiel, Alessandra Cervone, and Giuseppe Riccardi. Inquisitive mind: a conversational news companion. In Proceedings of the 1st International Conference on Conversational User Interfaces, pages 1–3, 2019

[8] Anushka Gupta, Diksha Chugh, Rahul Katarya, et al. Automated news summarization using transformers. In Sustainable Advanced Computing, pages 249–259. Springer, 2022.

[9] Gabriel Haas, Michael Rietzler, Matt Jones, and Enrico Rukzio. Keep it short: A comparison of voice assistants' response behavior. In CHI Conference on Human Factors in Computing Systems, pages 1–12, 2022.

[10] Hyunhoon Jung, Changhoon Oh, Gilhwan Hwang, Cindy Yoonjung Oh, Joonhwan Lee, and Bongwon Suh. Tell me more: Understanding user interaction of smart speaker news powered by conversational search. In Extended abstracts of the 2019 chi conference on human factors in computing systems, pages 1–6, 2019.

[11] Philippe Laban, John Canny, and Marti A Hearst. What's the latest? a question-driven news chatbot. arXiv preprint arXiv:2105.05392, 2021.

[12] Philippe Laban, Elicia Ye, Srujay Korlakunta, John Canny, and Marti Hearst. Newspod: Automatic and interactive news podcasts. In 27th International Conference on Intelligent User Interfaces, pages 691–706, 2022.

[13] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proceedings of the ACM on humancomputer interaction, 2(CSCW):1–31, 2018.

[14] Adam D Lelkes, Vinh Q Tran, and Cong Yu. Quiz-style question generation for news stories. In Proceedings of the Web Conference 2021, pages 2501–2511, 2021.

[15] Bang Liu, Haojie Wei, Di Niu, Haolan Chen, and Yancheng He. Asking questions the human way: Scalable question-answer generation from text corpus. In Proceedings of The Web Conference 2020, pages 2032–2043, 2020.

[16] Sara Irma Parnell, Stefanie Helene Klein, and Franziska Gaiser. Do we know and do we care? algorithms and attitude towards conversational user interfaces: Comparing chatbots and voice assistants. In Proceedings of the 4th Conference on Conversational User Interfaces, pages 1–6, 2022.

[17] Shruti Rao, Valeria Resendez, Abdallah El Ali, and Pablo Cesar. Ethical self-disclosing voice user interfaces for delivery of news. In Proceedings of the 4th Conference on Conversational User Interfaces, pages 1–4, 2022.

[18] Harshita Sahijwani, Jason Ingyu Choi, and Eugene Agichtein. Would you like to hear the news? investigating voice-based suggestions for conversational news recommendation. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, pages 437–441, 2020.

[19] Jing Wei, Tilman Dingler, and Vassilis Kostakos. Understanding user perceptions of proactive smart speakers. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(4):1–28, 2021.

Thank You !!!
Questions?
Vocal Assistant for Reading News

Jennifer Swaminathan

University of Neuchatel, Switzerland

SUPERVISOR:

Leonardo Angelini